https://eventos.utfpr.edu.br//sicite/sicite2018

# Identificação de disfonias Utilizando Redes Neurais Artificiais

# Dysphonias Identification Using Artificial Neural Network

**Aron Alexandre Martins Lima**
aron@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

**Rafael Martinelli de Araujo**
rafaelaraujo@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

**Fabio Augusto Guidotti dos Santos**
fsantos.1995@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

**Victor Hideki Yoshizumi**
yoshizumi@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

**Danilo Hernane Spatti**
spatti@icmc.usp.edu.br
Universidade de São Paulo, São Carlos, São Paulo, Brasil

**Maria Eugenia Dajer**
medajer@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

## RESUMO

As patologias laríngeas causam alterações no padrão vibratório das pregas vocais. Essas alterações produzem mudanças na qualidade vocal e podem tornar-se um problema significativo, principalmente para quem faz uso ocupacional e profissional da voz. Os métodos de avaliação de voz são em geral invasivos, possuem alto custo de implementação e causam desconforto ao paciente. O objetivo deste trabalho é a aplicação de procedimentos para pré-processamento de arquivos de áudios e aplicação da Transformada Wavelet Packet, separação de arquivos de treinamento e teste utilizando a Rede Neural Artificial Adaptive Resonance Theory e identificação de amostras disfônicas e saudáveis, utilizando a Rede Neural Artificial Perceptron Multicamadas, como método não invasivo de análise vocal.

**PALAVRAS-CHAVE:** Disfonia. Transformada Wavelet Packet. Perceptron Multicamadas. Adaptive Resonance Theory.

## ABSTRACT

Laryngeal pathologies cause changes in the vibratory pattern of the vocal folds. Producing changes in vocal quality which can become a significant problem, especially for those who make occupational and professional use of voice. The voice assessment methods are generally invasive, have a high implementation cost and cause discomfort to the patient. The objective of this work is the application of procedures for preprocessing audio files and application of the Wavelet Packet Transform, training and test files separation using the Artificial Neural Network Adaptive Resonance Theory and identification of dysphonic and healthy samples using the Artificial Network Neural Multilayer Perceptron, as non-invasive vocal analysis method.

**KEYWORDS:** Dysphonia. Wavelet Packet Transform. Multilayer Perceptron. Adaptive Resonance Theory.

**Recebido:** 31 ago. 2018.
**Aprovado:** 04 out. 2018.

## INTRODUÇÃO

As patologias na laringe causam alterações no padrão vibratório das pregas vocais (disfonias), ou seja, alterações na voz o que pode ser um problema para aqueles que à usam para trabalho. Behlau (2004) sugere que as disfonias podem ser classificadas em 3 grupos. As disfonias funcionais, causadas pelo mal-uso da voz, inadaptações da voz e alterações psicogênicas, ou seja, este é um distúrbio de comportamento vocal. As disfonias organofuncionais, lesões benignas de origem no comportamento vocal inadequado, ou seja, lesões formadas geralmente por disfonias funcionais sem tratamento, representadas neste trabalho por Nódulos. As disfonias orgânicas, são causadas por doenças neurológicas, inflamações, infecções na laringe, entre outros, representadas neste trabalho por Paralisia Unilateral de Prega Vocal.

Os procedimentos de avaliação vocal são classificados em invasivos e não invasivos. Os invasivos consistem na introdução de ferramentas de captura de imagens para analisar a estrutura e o padrão vibratório das pregas vocais (TSUJI et al, 2014), GONZÁLEZ, 2008). São de alto custo, causam desconforto ao paciente e dependem da experiência do examinador. Os não invasivos buscam por meio da análise do sinal de voz, padrões que caracterizem as disfônicas. Eles têm menor custo e causam menor desconforto ao paciente, além de possibilitar a criação de sistemas automatizados, para auxiliar na classificação de disfonias (SCALASSARA, 2009), (FERMINO et al,2016), (BARIZÃO et al, 2017).

Dessa forma, o objetivo do trabalho é identificar amostras disfônicas e saudáveis utilizando a RNA Perceptron Multicamadas (PMC). Para isso será realizado um pré-processamento de arquivos de áudios e aplicação da Transformada Wavelet Packet (TWP); a separação de arquivos de treinamento e teste será realizada com a RNA Adaptive Resonance Theory (ART).

## METODOLOGIA

Todos os procedimentos deste trabalho foram realizados utilizando a ferramenta de desenvolvimento MATLAB, nas dependências da Universidade Tecnológica Federal do Paraná campus Cornélio Procópio.

### BASE DE DADOS

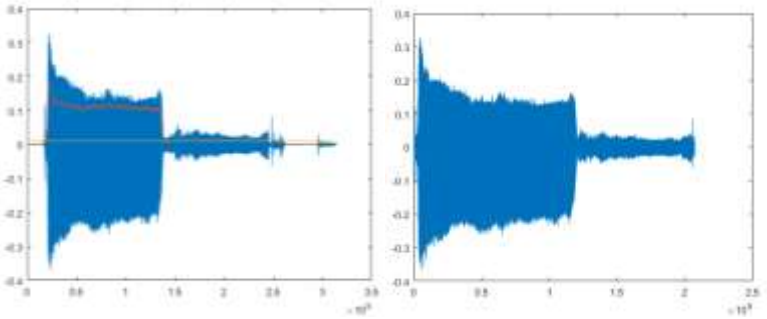
Os dados utilizados neste estudo são arquivos de áudios contendo a vogal /e/ sustentada, divididos em 2 grupos: vozes saudáveis, vozes com disfonia: Funcional, Organofuncional e Orgânica. Os áudios de Disfonia Funcional foram coletados na Universidade Federal de São Paulo (UNIFESP).(ZAMBON,2011, p,13) Os áudios de Disfonia Organofuncional, Disfonia Orgânica e Saudáveis, foram coletados no Ambulatório de Voz do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (HC-FMUSP) e cedido pelo Grupo de Engenharia Médica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (GPEMCNPq). Contando com, 21 áudios de disfonia organofuncional (Nódulos), 45 áudios de disfonia orgânica (paralisia unilateral de prega vocal) e 51 saudáveis (FERMINO,2017; PIMENTA, 2016).

## PRÉ-PROCESSAMENTO

Devido às características dos sinais e a diferente procedência dos dados foi preciso resolver os seguintes problemas: i) uniformidade de taxa de amostragem entre os áudios, ii) momentos de silêncio e iii) DC offset e artefatos nos sinais.

Para resolver a uniformidade das taxas de amostragens, foi necessário fazer uma reamostragem. Como o processo de *upsampling* gera amostras artificiais por interpolação, gerando possíveis ruídos, utilizou-se o *downsampling* com razão 2, que apenas elimina amostras. O segundo problema, o possível DC Offset foi resolvido utilizado uma função *detrend* do MATLAB que elimina possíveis desvios do sinal. Para remoção do silêncio foi utilizado um filtro de remoção das amostras, que consiste inicialmente na aplicação de um envelope RMS sobre o sinal, com apresenta a figura 5a em vermelho. Em seguida calculou-se o valor mínimo e médio do envelope. Para o cálculo do limitante de corte em amplitude, calculou-se a média das amostras do envelope que são maiores que o dobro do mínimo e menores que a metade da média das amostras totais, como pode ser visto na Figura 5a em amarelo. Finalmente eliminou-se todas as amostras com amplitude inferior ao limiar obtido, dessa forma o silêncio é removido com perda mínima de amostras. Obtendo-se o sinal da Figura 5b. Os artefatos como engasgos, tosses, vozes de terceiros, etc. foram removidos manualmente. Por último foi realizada uma normalização dos áudios por RMS.

Figura 5-(a) Sinal de voz, envelope RMS e limiar de corte ;(b) Sinal sem silêncio.



Fonte: Autoria própria (2018).

O seguinte procedimento realizado foi a remoção manual de artefatos. Artefatos como engasgos, tosses, vozes de terceiros, etc. Por último foi realizada uma normalização dos áudios por RMS.

## SEPARAÇÃO TREINAMENTO E TESTE

Para garantir a diversidade no conjunto de treinamento, utilizou-se as RNAs do tipo ART, utilizadas para agrupamento de dados, as quais foram implementadas durante este trabalho. Com estas foi possível, observar similaridades entre áudios de um mesmo grupo. Aplicou-se a TWP da família Daubechies 2, com 3 níveis de decomposição, janelamento de 2048 amostras e sobreposição (*overlap*) de 50%. A energia decimal obtida com a TWP, foi convertida em binária de 15 bits, com valor máximo igual a 100. Os valores obtidos, foram, portanto, utilizados com entradas da ART, onde os áudios que possuíam maior similaridade com os outros, foram selecionados para o conjunto de testes, resultando em 20% dos dados para teste e 80% para treinamento.

## EXTRAÇÃO DE CARACTERÍSTICA

Com os conjuntos de treinamento e teste separados, foram aplicados um conjunto de topologias de TWP, sendo elas das Famílias Daubechies 2 a 10, Symlets 2 a 10, e coiflets 1 a 5, níveis de decomposição 3, 4 e 5, com janelamento dos dados de 1024, 2048, 4096 e 8192 amostras e sobreposição (overlap) de 50%, 75%, 85% e 90%, totalizando 1104 topologias testadas, com objetivo de buscar qual delas possuem melhores resultados na identificação de disfonias. As informações extraídas da TWP foram a energia, a entropia de shannon e o logaritmo da energia. Após a extração de características a partir da TWP os vetores de saídas foram gerados e concatenados com os arquivos, onde o vetor de saída [1 0] representa os dados saudáveis e o vetor [0 1] representa os dados disfônicos.

## CLASSIFICAÇÃO

Para o processo de classificação foi utilizada a RNA do tipo PMC com confiabilidade de 98%, ou seja, valores que a rede apontar acima de 0,98 são considerados iguais a 1 e valores abaixo de 0,02 são considerados 0, o restante é considerado incerteza. Com 3 camadas intermediárias com 10, 9 e 7 neurônios respectivamente, taxa de aprendizagem igual a 0,3, função de ativação tangente hiperbólica, algoritmo de treinamento Backpropagation Levenberg-Marquardt.

## RESULTADOS E DISCUSSÃO

A partir dos procedimentos descritos no tópico anterior, obteve-se os arquivos de teste. Onde a quantidade de amostras utilizadas no teste, correspondem a 20% do total de amostras. Com os arquivos de teste e treinamento separados foi possível testar todas as topologias de Transformadas Wavelets citadas. Foram realizados 3 diferentes treinamentos para cada topologia, a taxa de acertos para cada uma das classes (Saudáveis e Disfônicos) das 5 melhores topologias, estão apresentados na Tabela 1.

Tabela 1 – Melhores topologias TWP.

| **Symlet 3, nível 5, janela 8192, overlap 90%** | **Percentual** |
|---|---|
| Saudáveis | 86,78% |
| Disfônicos | 91,76% |
| **Symlet 3, nível 3, janela 8192, overlap 90%** | **Percentual** |
| Saudáveis | 88,13% |
| Disfônicos | 89,40% |
| **Symlet 2, nível 3, janela 1024, overlap 90%** | **Percentual** |
| Saudáveis | 93,75% |
| Disfônicos | 83,65% |
| **Symlet 2, nível 3, janela 1024, overlap 50%** | **Percentual** |
| Saudáveis | 97,64% |
| Disfônicos | 82,78% |
| **Daubechies 2, nível 5, janela 8192, overlap 85%** | **Percentual** |
| Saudáveis | 89,31% |
| Disfônicos | 90,24% |

Fonte: Autoria Própria (2018).

Como pode-se observar no quadro a topologia da TWP que apresentou melhor resultado total foi a família Daubechies 2, com 5 níveis de decomposição, janelamento de 8192 amostras e sobreposição de 85%, onde foram utilizadas 2215 amostras para teste, dentre elas 739 saudáveis e 1476 disfônicas. A tabela 2, apresenta a matriz confusão do melhor resultado obtido:

Tabela 2 – Matriz confusão melhor topologia TWP.

| | | Saudáveis | Disfônicos | Incerteza |
|---|---|---|---|---|
| Desejado | Saudáveis | 89,31% | 2,16% | 8,53% |
| | Disfônicos | 8,94% | 90,24% | 0,82% |

Fonte: Autoria Própria (2018).

Desta forma pode-se observar, que a partir da metodologia sugerida, apresentou-se bom resultado com total de acertos de 89,93% na classificação de disfonias com confiabilidade de 98%, mesmo sem a otimização dos parâmetros da Perceptron Multicamadas.

## CONCLUSÃO

Diante das necessidades observadas, foi possível elaborar um sistema utilizando conceitos computacionais, como pré-processamento dos áudios de modo a tratar diferenças nas taxas de amostragem, DC offset, momentos de silêncio e artefatos externos, o tratamento de sinais utilizando a TWP, as RNAS do tipo ART e PMC resultando em um sistema eficaz em classificação de padrões auxiliando na triagem de pacientes disfônicos. Com isso, o resultado obtido foi uma classificação com taxa de acertos na ordem de 90%, com confiabilidade de 98%, o que indica uma classificação eficiente. Diante desses resultados, abre-se então espaços para estudos na aplicação de sistemas inteligentes na classificação de disfonias. Futuramente, espera-se otimizar os parâmetros da PMC, visando melhores resultados, bem como a segregação dos arquivos disfônicos.

## REFERÊNCIAS

BARIZÃO, A. H. Estimação do Grau de Parâmetros Subjetivos Vocais Aplicando Redes Neurais Artificiais. 2017. 65 p. Trabalho de Conclusão de Curso - Engenharia Elétrica - Universidade Tecnológica Federal do Paraná.

BEHLAU, M. A voz: O livro do especialista. Vol. I. Rio de Janeiro, RJ: Revinter, 2004.

FERMINO, M. A. et al Classificação de Distúrbios Vocais Aplicando Redes Neurais Artificiais e Transformada Wavelet Packet. 12th IEEE/IAS International Conference on Industry Applications – INDUSCON. Curitiba – PR, Brazil (2016).

GONZÁLEZ, I. V. Videolaringoestroboscopia: una técnica para visualizar las cuerdas vocales. Estudios de fonética experimental, no. 17: p. 418-432, 2008.

PIMENTA, R. A. Uso da Avaliação Multidimensional da Voz na Caracterização Vocal de Pacientes com Paralisia Unilateral de Pregas Vocais. (Tese de Doutorado em Ciências)- Programa de Pós-Graduação Interunidades Bioengenharia, Escola de Engenharia de São Carlos, São Carlos, 2016.

SCALASSARA, P. R. Utilização de Medidas de Previsibilidade em Sinais de Voz para Discriminação de Patologias da Laringe. (Doutorado em engenharia Elétrica) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2009.

TSUJI, D. H. et al. Improvement of Vocal Pathologies Diagnosis Using High-Speed Videolaryngoscopy. International Archives of Otorhinolaryngology, Rio de Janeiro, v. 18, n. 3, p. 294-302, abr. 2014.

ZAMBON, F. C. Estratégias de Enfrentamento em Professores com Queixa de Voz . 2011. 13-19 p. (Dissertação Mestrado em Ciências)- Escola Paulista de Medicina, Universidade de São Paulo, São Paulo, 2011.

## AGRADECIMENTOS

Agradecemos a Fabiana Zambon e ao Grupo de Engenharia Médica do Conselho Nacional de Desenvolvimento Científico e Tecnológico, por disponibilizar as bases de dados que tornaram este estudo possível.